

2023 S.T. Yau High School Science Award (Asia)

Research Report

The Team

Registration Number: Comp-169

Name of team number: Lim Li Xin Jed
School: NUS High School of Mathematics and Science
Country: Singapore, Singapore

Name of team number: Qiu Ziming
School: NUS High School of Mathematics and Science
Country: Singapore, Singapore

Title of Research Report

μ -Net: ConvNext-Based U-Nets for Cosmic Muon Tomography

Date:

August 15, 2023

Abstract

Muon scattering tomography is an imaging technique that utilizes muons, typically originating from cosmic rays to image the interiors of objects. Since muons are highly penetrating, muon tomography can be used to investigate the internal composition of dense materials such as geological formations or archaeological structures. However due to the low flux of cosmic ray muons at sea-level and the highly complex interactions that muons display when travelling through matter, existing reconstruction algorithms often suffer from low resolution and high noise. In this work, we develop a novel two-stage deep learning algorithm, μ -Net, consisting of an MLP to predict the muon trajectory and a ConvNeXt-based U-Net to convert the scattering points into voxels. μ -Net is trained on synthetic data generated by the Geant4 simulation package and we show that it outperforms existing reconstruction methods for muon tomography. It achieves a state-of-the-art performance of 17 PSNR at a dosage of 1024 muons, outperforming traditional reconstruction algorithms such as the point of closest approach algorithm and maximum likelihood and expectation maximisation algorithm. Furthermore, we find that our method is robust to various corruptions such as inaccuracies in the muon momentum or a limited detector resolution. We hope that this research will spark further investigations into the potential of deep learning to revolutionise this field. Our model code is at <https://github.com/jedlimlx/Muon-Tomography-AI> and our data generation code is at <https://github.com/jedlimlx/Muons-Data-Generation>. Our dataset can be found at <https://www.kaggle.com/datasets/tomandjerry2005/muons-scattering-dataset>.

Keywords: Muon Tomography, Geant4, U-Net, ConvNeXt, CNN

Acknowledgements:

- We would like to sincerely thank Kannan Vishal for writing an initial simulation for the muon scattering.

Commitments on Academic Honesty and Integrity


We hereby declare that we

1. are fully committed to the principle of honesty, integrity and fair play throughout the competition.
2. actually perform the research work ourselves and thus truly understand the content of the work.
3. observe the common standard of academic integrity adopted by most journals and degree theses.
4. have declared all the assistance and contribution we have received from any personnel, agency, institution, etc. for the research work.
5. undertake to avoid getting in touch with assessment panel members in a way that may lead to direct or indirect conflict of interest.
6. undertake to avoid any interaction with assessment panel members that would undermine the neutrality of the panel member and fairness of the assessment process.
7. observe the safety regulations of the laboratory(ies) where the we conduct the experiment(s), if applicable.
8. observe all rules and regulations of the competition.
9. agree that the decision of YHSA(Asia) is final in all matters related to the competition.


We understand and agree that failure to honour the above commitments may lead to disqualification from the competition and/or removal of reward, if applicable; that any unethical deeds, if found, will be disclosed to the school principal of team member(s) and relevant parties if deemed necessary; and that the decision of YHSA(Asia) is final and no appeal will be accepted.

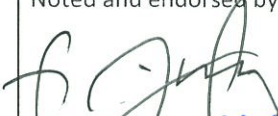
(Signatures of full team below)


 Name of team member: *Lim Li Xin* *Jed*


 Name of team member: *Qiu Ziming*

Name of team member:

Name of supervising teacher:
Ng Chee Loon 

Noted and endorsed by  (signature) Mr Goh Hock Leong Deputy Principal (Academic) NUS High School of Math & Science Name of school principal:
--

Contents

Abstract	ii
Acknowledgements	iii
Commitments on Academic Honesty and Integrity	iv
1 Introduction	1
2 Preliminaries	2
2.1 Physical Background	2
2.2 Problem Statement	2
2.3 Motivation	3
3 Related Work	3
3.1 Deep Learning Methods	3
3.2 Traditional Algorithms	4
4 Methods	4
5 Training Techniques.	5
6 Experiments	6
6.1 Experimental Setup	6
6.2 Ablations	7
6.3 Comparison with Traditional Algorithms	8
7 Discussion	10
7.1 Applications	10
7.2 Future Work	10
8 Conclusion	12
Appendices	17
Appendix A Proofs	17
A.1 Arbitrarily Large Resolution	17
A.2 Arbitrarily Large Point Size	17

Appendix B Raw Results

19

1 Introduction

Muon tomography is an imaging technique that utilizes muons, typically originating from cosmic rays, to image the interiors of objects. By leveraging the fact that muons are highly penetrating particles, muon tomography offers a non-invasive and non-destructive means of investigating the internal composition of dense materials. Through tracking and analysis of muon trajectories and energies, this enables the accurate reconstruction of internal structures and features. Consequently, muon tomography has emerged as a tool with wide-ranging applications in fields such as geophysics [1], civil engineering [2], and archaeology [3, 4].

There are 2 main types of muon tomography - muon scattering tomography and muon absorption tomography. Muon absorption tomography relies on the measurement of the attenuation of muon flux as it traverses a material. In muon scattering tomography, the directions of the scattered muons are also considered. We focus on muon scattering tomography. They both make use of muon detectors to record the flux of cosmic muons passing through an object. The difference between the typical cosmic muon flux and the detected flux of muons provides information about the interior of the objects.

Due to cosmic rays entering the Earth's atmosphere, there is no need for a specialised muon source unlike other types of tomography. However, there are many other challenges in this task. First, the flux of cosmic muons at sea-level is very low [5]. In order to produce a decent reconstruction, data has to be collected for a long period of time. Furthermore, unlike other types of tomography such as x-ray computed tomography, muons will scatter off atomic nuclei. This makes the forward operator highly nonlinear. In contrast, in computed tomography, x-rays only attenuate and hence, the forward operator is the linear radon transform. Furthermore, due to limitations of current day muon detectors, the momentum of muons which significantly affects the muon scattering angle is not known. At best, we will have an estimate of the momentum \hat{p} with a significant amount of uncertainty.

Several methods have been developed to tackle this problem. The first algorithm developed was the Point of Closest Approach (PoCA) algorithm [6] which assumes that the muons only scatter once at the point of closest approach of its inward and outward trajectory. The Maximum Likelihood and Expectation Maximisation (MLEM) algorithm [7] improves on PoCA and iteratively optimises the reconstruction to maximise the likelihood of producing a given scattering outcome. Other algorithms have also been developed such as most probable trajectory (MPT) [8], maximum a posteriori (MAP) [9], most likely path [10, 11, 12], scattering density estimation (SDE) [13] and angle statistics reconstruction (ASR) [14]. Other methods have also been developed for low dosages such as the binned clustering algorithm [15] and a method based on density clustering [16].

However, no one has attempted to make use of developments in deep learning to solve this ill-posed problem. Due to the abundance of data which can be obtained from simulations from software such as Geant4 [17] and the non-linearity of the problem, deep learning methods are well-suited for this task. They have been previously applied to other inverse problems such as computed tomography [18].

In this work, we develop a novel two-stage deep learning algorithm, μ -Net, for cosmic muon scattering tomography, based on using the point of closest approach (PoCA) algorithm and the U-Net architecture proposed by Ronneberger et al. [19]. Instead of using the more traditional Residual Blocks, we make use of ConvNeXt Blocks [20] which result in better performance than Residual Blocks at a lower computational

cost. We find that our model significantly out-performs traditional reconstruction algorithms and achieves state-of-the-art performance.

2 Preliminaries

2.1 Physical Background

Muon scattering relies primarily on modelling the scattering interaction between muons and matter. Rossi et al. [21] developed a scattering theory for charged particles and found that charged particles travelling through a plate of thickness x that undergo coulomb scaling have scattering angles and lateral displacements that follow a Gaussian distribution with mean, $\mu = 0$ and variance,

$$\sigma^2 = \frac{E_s^2}{2p^2v^2} \frac{x}{L_{rad}} \quad (1)$$

where $E_s = 21$ MeV, λ is the radiation length of the material, x is the thickness of the plate and p and v are the momentum and velocity of the muon respectively.

From this formula, we can define the parameter of interest, the scattering density λ .

$$\lambda = \frac{\sigma^2}{x} = \frac{15MeV}{p^2v^2} \frac{1}{L_{rad}} \quad (2)$$

With this, the task of muon tomography is to find the distribution of λ within the object through looking at positions and directions of the incoming and outgoing muons.

2.2 Problem Statement

Before proceeding with a literature review and the description of the method, let us formalize the muon reconstruction problem. We shall follow a similar notation to Schultz et al. [7].

Let the object of interest be defined by its scattering density,

$$\lambda(x, y, z) = \left(\frac{13.6}{p_0} \right)^2 \frac{1}{L_{rad}(x, y, z)} \quad (3)$$

where $L_{rad}(x, y, z)$ is the radiation length at each point within the object.

We can represent the scattering density in terms of some basis functions $\phi_j(x, y, z)$ such that

$$\lambda(x, y, z) = \sum_j \alpha_j \phi_j(x, y, z) \quad (4)$$

where $\alpha = [\alpha_j]$ are the coefficients for the basis functions.

Suppose the muon detections, \mathbf{y} follow a distribution D parameterized by the scattering density of the object, i.e.

$$\mathbf{Y} \sim D(\alpha) \quad (5)$$

Therefore, given a sample of n muon detections $Y_n = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$ we wish to construct a point estimate $\mathbf{a}(Y_n)$ of $\boldsymbol{\alpha}$, which is approximated by the deep neural network $f_{\boldsymbol{\theta}}(Y_n)$ parameterized by $\boldsymbol{\theta}$.

In this paper, we will take p_0 to be 13.6 MeV, so that we directly regress the reciprocal of the radiation length $\frac{1}{L_{rad}(x,y,z)}$ in units of cm^{-1} .

2.3 Motivation

Some key features of the muon reconstruction problem are immediately apparent from the problem statement, which help us design our model architecture:

- The model should be permutation-invariant, i.e. the order in which muons are inputted into the model should not change the output.
- The model should accept any number of input muons.
- The model should be able to make use of paired input and output muon detections i.e. the model needs to know which input muon corresponds to which output.
- The model should be able to take advantage of the 3D spatial structure of the target output.

This seems to suggest making use of Transformers [22] with the positional encoding removed. However, the dot product attention module used in Transformers has a large time-complexity of $O(N^2)$ where N is the dosage. Furthermore, the transformer will be unable to take advantage of the spatial structure of the output. Hence, we will make use of the PoCA algorithm and the U-Net architecture [19] to create a two-stage algorithm.

3 Related Work

3.1 Deep Learning Methods

Point Clouds. From the key features of the muon tomography problem, we see that it is highly similar to point cloud problems, which also take in permutation invariant data (a set of points), but the points themselves also exhibit some 3D structure. Deep learning methods on point cloud data can generally be classified into 2 main categories, neural networks which operates directly on the points, and neural networks which operate on a voxelized representation of the data [23]. In our work, we chose the latter approach since the former tends to be slightly worse at capturing spatial structure, and our desired output is also in the form of voxels.

U-Net. The U-Net was first proposed as an medical image segmentation model [19]. It consists of a downward branch, where the image is downsampled and an upward branch, where the image is upsampled. Skip connections are also used between the downward and upward branches to allow high resolution features, which may be removed during downsampling, to be retained. U-Nets have also been extended to process 3D data [24, 25], by replacing the standard 2D convolution operations with 3D convolutions.

ConvNext. ConvNext is a recent iteration of the family of convolutional neural networks [20]. It was proposed as an improvement of the ResNet [26] by incorporating methods found in Vision Transformer architectures, most notably the SWIN transformer [27], and has been shown to obtain competitive performance against Transformer-based methods but with a lower parameter count. In our work, we will use a modified 3D ConvNext block to form our U-Net.

3.2 Traditional Algorithms

Point of Closest Approach. The Point of Closest Approach (PoCA) method is a commonly used method for muon scattering tomography, first proposed by Schultz et al. [6]. It assumes that the muon is scattered once by an object at the point of closest approach between the input and output trajectories. However, this fails to take into account the fact that in many cases, muons may scatter multiple times. As a result, there will be predictions that muons scatter at points where there is actually no object as illustrated in Figure 1. Furthermore, some information is lost as some muons may have a point of closest approach outside of the object space.

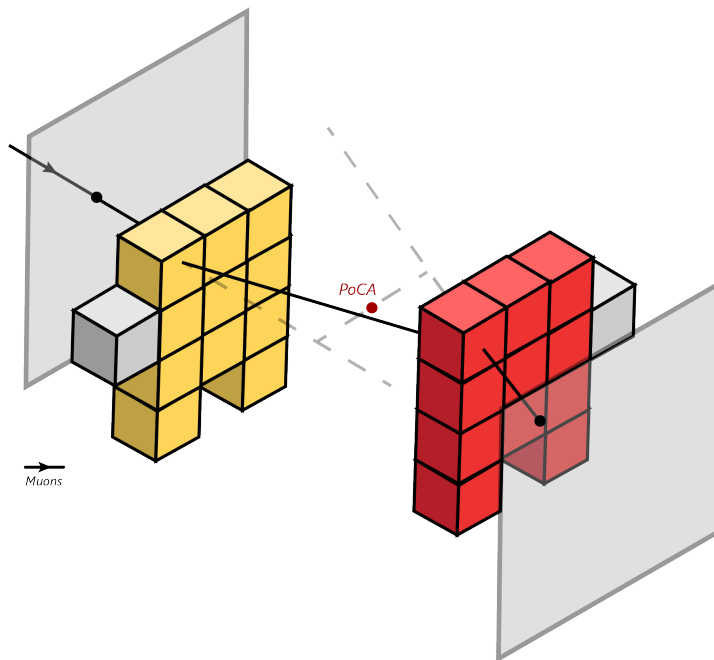


Figure 1: A diagram illustrating the point of closest approach algorithm. Since the muon scatters more than once, the computed PoCA is not within any object.

Maximum Likelihood Expectation Maximization. The maximum likelihood expectation maximization (MLEM) algorithm is a statistical reconstruction method proposed by Schultz et al. [7]. It makes use of the statistical distribution of muon scattering to frame muon reconstruction as a maximum likelihood problem. An iterative expectation maximization algorithm is then used to find the scattering densities within the object that is most likely to result in the observed data collected by the muon detectors.

Existing literature [28] have found that MLEM has better qualitative performance than direct allocation to PoCA. However, in our experiments, we observed that MLEM has significantly lower performance (PSNR of 13.54 for a dosage of 2048 muons against 14.40 for PoCA). We hypothesize that this is due to the lower muon dosages we used in our experiments, which makes it difficult to apply statistical methods to the reconstruction. Due to the lower performance and higher computational cost of MLEM, we will be comparing our model against PoCA for most of our experiments.

4 Methods

Scatter Operation. Our goal is to convert a set of muon detections into a reconstruction output. One simple way to do this is to first convert the muon detections into some 3D representation, which can then be fed into a U-Net.

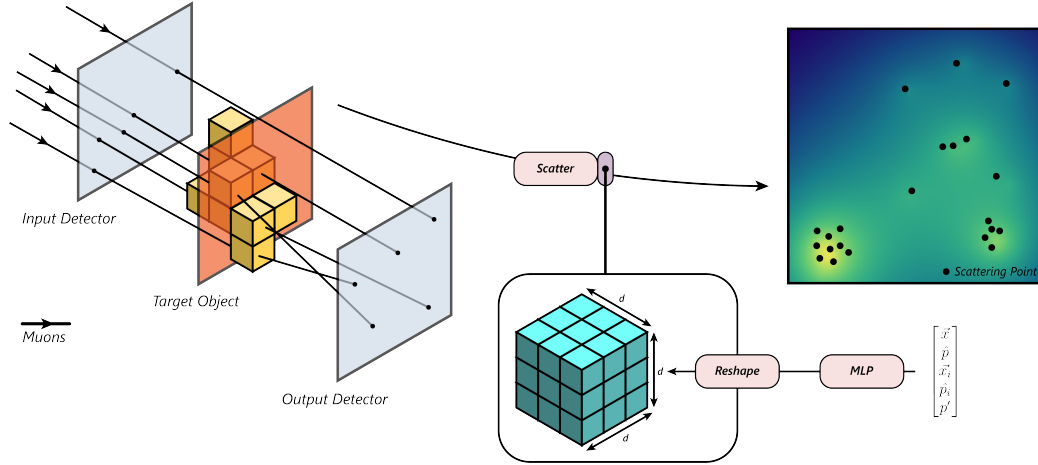


Figure 2: A diagram of the first stage of μ -Net. The muon features (initial position, initial momentum, etc.) are passed through an MLP and placed at PoCA scattering points within a 3D volume. If they overlap, the average is taken.

To achieve this, we first apply an MLP on the muon’s input parameters such as the initial momentum, initial position, final position and final momentum. The output features are reshaped into a $d \times d \times d \times c$ block. We call d the point size and c is the number of channels. Now, using the PoCA algorithm, we find the muon’s point of closest approach and scatter the output features into the voxels near the PoCA point. That is, we find the indices of the $d \times d \times d$ block of voxels that are closest to the PoCA point and set the voxel value at those points to the MLP output. If there is overlap in the scattering voxels, the average is taken. We choose the PoCA point as the information about the muons that scatter in a given area is the primary information that is needed to help decide what the scattering density of that area is.

A benefit of this approach is the scaling of the method’s runtime with dosage. Since the scattering of the muon detections into the voxelized volume and the computation of the scattering points is very fast, the processing time of the U-Net is much longer. As a result, our method effectively runs in constant time. With a dosage of 1024 and a batch size of 8, the model has an inference time of 171 ms. With a dosage of 16384 and a batch size of 8, the model runs at 246 ms.

U-Net. Now, we make use of the U-Net [19] to process the voxelised volume matrix from the first stage. In our U-Net, instead of using the more traditional Residual Blocks [26], we make use of ConvNeXt Blocks [20] which result in better performance than Residual Blocks at a lower computational cost. Each layer of the U-Net contains multiple such ConvNeXt blocks. Downsampling is done using Layer Normalisation followed by a convolutional layer of strides = 2. Upsampling is done by first applying a pointwise convolution and layer normalisation before using nearest neighbour upsampling.

5 Training Techniques.

We train our model using the Adam optimizer with a learning rate of 1×10^{-3} . We make use of progressive training and first train the model on a dosage of 1024 muons for 15 epochs. Then, we initialise the model with those weights and train it on 2048 muons for 15 epochs, and so on.

Universal Approximation. We have shown that μ -Net is an universal function approximator for continuous set functions given the model is large enough and either of the following conditions are met:

- The resolution of the reconstructed volume is sufficiently large such that there is no overlap in the reconstructed points **or**,
- the resolution of the reconstructed volume is finite but the point size is large enough to cover the entire reconstructed volume.

Formally, we can define $\chi = \{S : S \subseteq \mathbb{R}^m \text{ and } |S| = n\}$ and $f : \chi \rightarrow \mathbb{R}$ is a continuous set function w.r.t the Hausdorff distance $d_H(\cdot, \cdot)$. Our theorem proves that f can be arbitrarily approximated by our model if the resolution is sufficiently high, or if the resolution is fixed but the point size d is the same as the resolution.

Theorem 1. *Suppose $f : \chi \rightarrow \mathbb{R}^p$ is a continuous set function w.r.t $d_H(\cdot, \cdot)$, such that for all $\epsilon > 0$, there exists some configuration of the model parameters θ for sufficiently large p **or** $\phi(\eta(x_i)) = J_{p \times d}$ (i.e. the indicator function maps to every point), such that for any $S \in \chi$,*¹

$$\left| f(S) - \gamma_\theta \left(\frac{\sum_{x_i \in S} \{\phi(\eta(x_i)) \cdot h_\theta(x_i)\}}{\sum_{x_i \in S} \{\phi(\eta(x_i)) \cdot J_{d \times c}\}} \right) \right| < \epsilon$$

where $\gamma_\theta : \mathbb{R}^{p \times c} \rightarrow \mathbb{R}^p$ is any continuous function, $h_\theta : \mathbb{R}^m \rightarrow \mathbb{R}^{d \times c}$ is any continuous function, $\eta : \mathbb{R}^m \rightarrow \mathbb{R}$, $\phi : \mathbb{R}^m \rightarrow \mathbb{R}^{p \times d}$ and $J_{d \times c}$ is the ones matrix of shape (d, c) . η represents the PoCA function that generates a scattering point from the muon detection. ϕ is an indicator function for a set of intervals of length d derived from its input. The indicator function for each of these intervals is placed along one row in the last dimension. γ_θ and h_θ can be taken to be any continuous function due to the universal approximation theorem for CNNs and MLPs.

A brief proof is provided in the Appendix. We also note this theorem generalises to all dimensions easily by replacing p with $r \times r$, $r \times r \times r$, etc.

The limitations of the theorem to arbitrarily large resolutions or point sizes comes from the PoCA and scatter operations (i.e. $\phi(\eta(x_i))$). The nature of this function depends on the interactions between the muons and the object and is very difficult to analyze due to the complex interactions that muons exhibit with matter.

However, despite these limitations, we hypothesise that our model performs well with a fixed resolution and point size since most of the information about the muon scattering is contained near the the scattering point.

6 Experiments

6.1 Experimental Setup

Our data is generated using CERN's Geant4 simulation software [17]. To generate 3D objects to be analysed by our system, we make use of fractal noise to generate objects of various shapes. The material of the object is randomly chosen from a set list of materials of different radiation lengths.

The geometry of the system can be found in Figure 3. The target object is contained within a cube of side length 1 m. The input and output detectors are squares of side length 2 m. They are separated from the object by a distance of 0.5 m.

For the muon beam, we use a beam with a \cos^2 angular distribution and a power law distribution, in accordance with characterised values of the cosmic muon flux [29]. Muons that are calculated not to hit the detector are killed at the start of the simulation to ensure the simulation runs at a reasonable speed for data generation.

¹For this division, we assume that the division is element-wise and if division by 0 occurs the value will just be 0.

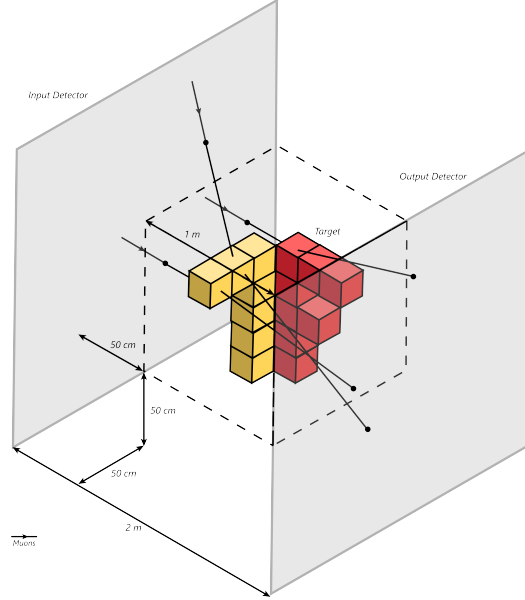


Figure 3: A scale diagram (except the voxels) illustrating the simulation setup within Geant4.

Table 1: Results of the model at difference dosages for various point sizes. The inference times are evaluated on 2 T4 GPUs with a batch size of 8. The best results are bolded.

Point Size	Dosage	Time↓	MSE↓	MAE↓	PSNR↑
1	1024	171 ms	0.2697	0.2417	16.3562
3	1024	171 ms	0.2674	0.2245	16.4496
5	1024	179 ms	0.2685	0.2428	16.3598
1	2048	175 ms	0.2217	0.2151	17.2552
3	2048	175 ms	0.2009	0.1883	17.7133
5	2048	198 ms	0.1992	0.1953	17.7513
1	4096	180 ms	0.1716	0.1691	18.4065
3	4096	187 ms	0.1701	0.1698	18.4348
5	4096	250 ms	0.1731	0.1742	18.3674
1	8192	201 ms	0.1428	0.1488	19.2274
3	8192	220 ms	0.1451	0.1486	19.1594
5	8192	364 ms	0.1433	0.1463	19.2040
1	16384	246 ms	0.1284	0.1340	19.7184
3	16384	296 ms	0.1321	0.1269	19.6145
5	16384	-	-	-	-

We use 3520 samples for the training set, 240 samples for the validation set and 1116 samples for the test set.

Our model is implemented using Tensorflow and trained using 2 T4 GPUs.

6.2 Ablations

Point Size. First, we vary the point size at various dosages to see its impact on the model’s performance.

We observe that for lower dosages, larger point sizes of 3 and 5 give the best performance. On the other hand, at higher dosages, smaller point sizes of 1 give better results. This is not unexpected. As the dosage increases, the number of scattering points also increases. This means that there is less of a need to have large

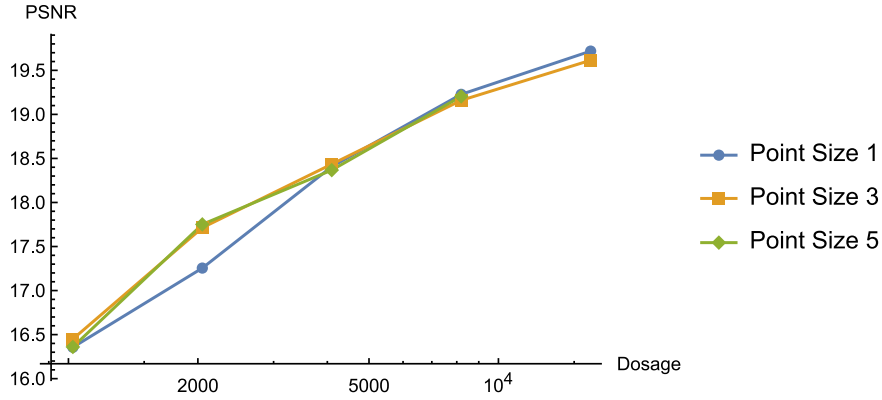


Figure 4: The results of Table 1 presented in a graph.

point sizes to fully "fill" up the 3D volume. As such, we will use a point size of 3 for dosages lower than 4096 and a point size of 1 for dosages higher than 4096.

Block Types. We also tried using a Residual Block to see how that would impact the performance of the model. As can be seen in Table 2, the Residual Block results in a significant drop in performance, hence, justifying our usage of ConvNeXt. Furthermore, we also see that the ConvNeXt-based model runs much faster as compared to the ResNet-based model, showing that our approach is also efficient.

Table 2: Results of the model when using ConvNeXt blocks and Residual Blocks. The best results are bolded.

Block Type	Dosage	Time↓	MSE↓	MAE↓	PSNR↑
ConvNeXt	1024	171 ms	0.2674	0.2245	16.4496
Residual	1024	412 ms	0.2754	0.2682	16.3152
ConvNeXt	2048	175 ms	0.2009	0.1883	17.7133
Residual	2048	479 ms	0.2402	0.2144	16.8412
ConvNeXt	4096	187 ms	0.1701	0.1698	18.4348
Residual	4096	467 ms	0.1965	0.1907	17.7513
ConvNeXt	8192	220 ms	0.1939	0.1486	19.1594
Residual	8192	465 ms	0.1939	0.2140	17.6638
ConvNeXt	16384	296 ms	0.1321	0.1269	19.6145
Residual	16384	440 ms	0.1575	0.1738	18.6310

6.3 Comparison with Traditional Algorithms

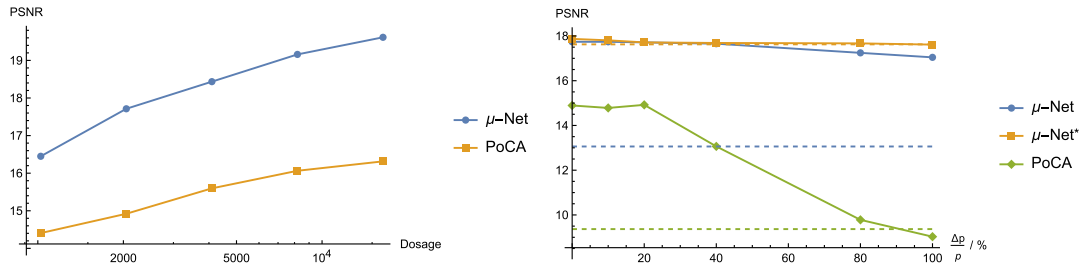


Figure 5: The PSNR of various methods plotted against (left) dosage levels and (right) percentage error in the momentum estimate. The lines represent the performance of the model when no momentum estimate is provided. μ -Net* represents the model's performance when it is fine-tuned on the new data.

Dosage. We vary the dosage of muons from 1024 to 16384. The results are shown in Table 3 and Figure 5 (left). We see that as the dosage increases, the performance of the models increase as well. Furthermore, it is clear that our model, μ -Net, significantly outperforms the the traditional PoCA algorithm.

Momentum Estimate. We also look at how varying levels of error in the momentum will affect predictions in Figure 5 (right). We again find that our model significantly outperforms the traditional PoCA algorithm. Furthermore, we find that after fine-tuning, the model's performance stays relatively constant as the error in the momentum increases, indicating our model is robust.

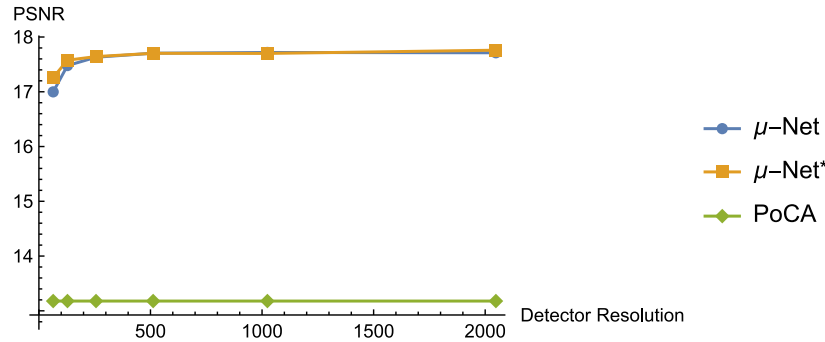


Figure 6: The PSNR of various methods plotted against various detector resolutions. μ -Net* represents the model's performance when it is fine-tuned on the new data.

Detector Resolution. Finally, we look at how the model's performance changes with the detector resolution in Figure 6. Again, we find that our model significantly outperforms the PoCA algorithm. Furthermore, we find that our model performs well at a variety of resolutions, showing that it is very robust.

Visual Comparison. Now, we visually compare the reconstructions of PoCA and MLEM. These reconstructions are shown in Figure 7 and 8. A dosage of 16384 muons is used. The MLEM reconstruction is significantly worse since it had to be done using a lower resolution, because the algorithm requires the muon tracks to pass through every voxel, which is not possible at higher resolutions.

First, we see that μ -Net provides superior reconstruction quality to PoCA. Furthermore, we can also see clearly that μ -Net is able to distinguish different materials by their radiation length and accurately reconstruct the approximate shape of objects. Nevertheless, we still observe significant levels of blurring.

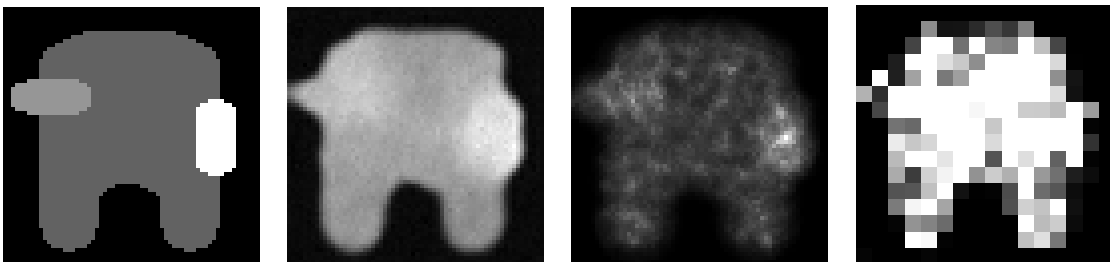


Figure 7: This is a cross-section of an object made of 3 different materials. The ability of the model to reconstruct the approximate shapes and differentiate materials is shown clearly. However, there is still a significant amount of blurring. (left) ground truth, (middle-left) μ -Net, (middle-right) PoCA (right) MLEM. MLEM was performed at a much lower resolution because it requires muon tracks to intersect with every voxel.

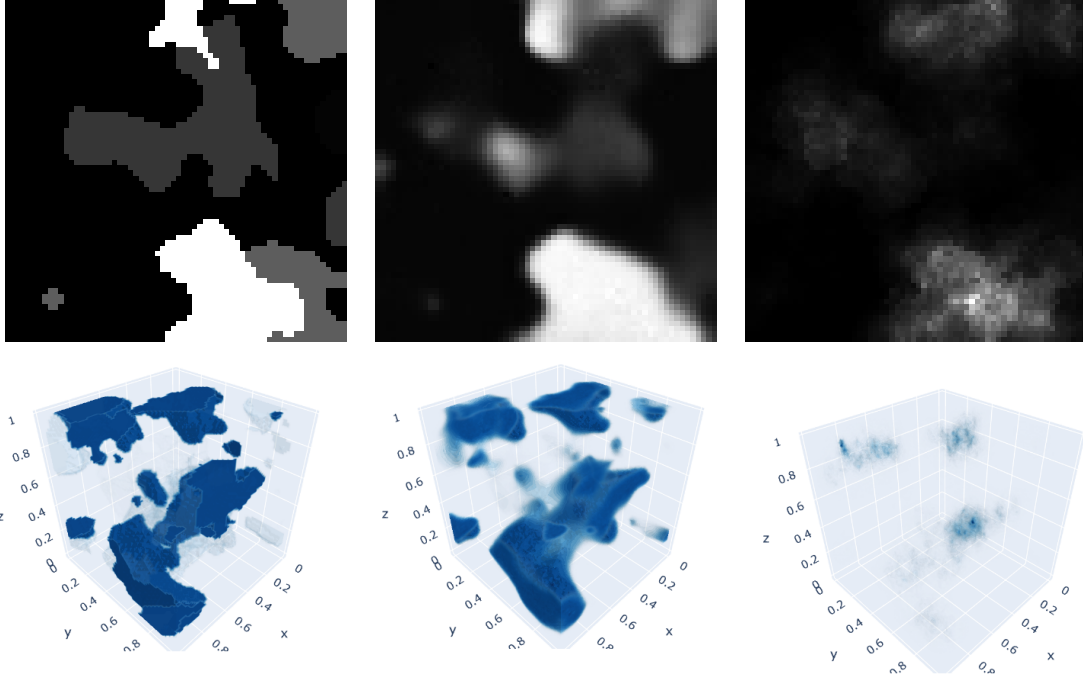


Figure 8: This is a 3D plot and a cross-section of one of the imaging targets from the testing set. The ability of the model to reconstruct the approximate shapes and differentiate materials is shown clearly. However, there is still a significant amount of blurring. (left) ground truth, (middle) μ -Net, (right) PoCA

7 Discussion

7.1 Applications

Nuclear Non-proliferation. Muon tomography has been applied in detecting the presence of high-Z materials such as radioactive materials since these materials result in a large amount of scattering. We test our model on such cases by generating a dataset containing high-Z objects placed within a container made of low-Z. We fine-tune our model on this dataset for 10 epochs.

As seen in Figure 9, our model is able to accurately reconstruct the positions and sizes of these objects. However, in some cases, the reconstruction is not as good. This is most likely caused by a thicker wall with a lower radiation length shielding the materials within the box.

Archaeology. In addition, muon tomography also has numerous applications in archaeology. In particular, the ability of muons to easily penetrate thick layers of material such as rock, enables their use to image the interiors of large structures. For instance, researchers have used muon tomography to discover secret chambers within Khufu’s Pyramid [4, 30], image underground cavities in Mount Echia [31, 32] and discover a secret ancient Greek burial chamber in the centre of Naples [33]. In cases where the input detections are available, our model can be used to significantly improve the accuracy of the reconstructions.

7.2 Future Work

No Input Muon Information. One limitation of the model is that it depends on the presence of information about the input muons because of its use of the PoCA algorithm. However, in some cases, such as in some archaeological applications, the direction of the input muon is unavailable.

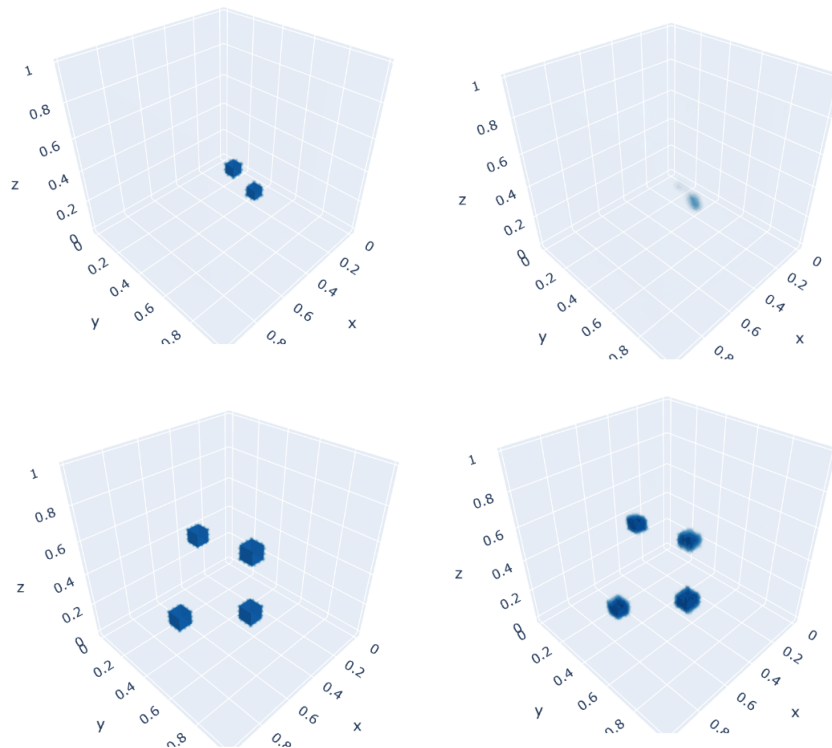


Figure 9: The model’s predictions for a high-Z object placed within a container made of low-Z material. The container is not visible as its radiation length is very large. The inverse of the radiation length is being plotted. (left) ground truth, (right) prediction

End-to-end Training. Another possibility for improvement is to develop an approach which consists of a single-stage and is end-to-end trainable. Currently, μ -Net makes use of the PoCA algorithm to compute the location of scattering points. This means that the model may still suffer from some limitations that PoCA faces such as false hotspots in the final detection when muons scatter two or more times.

Perception Losses. Another possible improvement would be the usage of perceptual losses [34]. These losses pass the reconstruction through an image classification model, pretrained on ImageNet, and compare feature vectors of the reconstruction and ground truth. They have seen success when training deep learning models for computed tomography [35]. These losses have an advantage over traditional loss functions such as MSE since they measure how the human eye would perceive the reconstruction. As a result, using this loss function can help to reduce blurring within the image, which we observe to be a significant issue with the current model. However, a major difficulty in using perception losses is that new classification models will have to be trained on 3D voxels, since existing models will likely not adapt well.

Trajectory Prediction. As shown in Figure 1, when muons scatter more than once, their PoCA will be outside of the bounds of any object. This is a significant limitation of PoCA and sometimes results in false hotspots of scattering density in the final prediction. One solution to this could be to attempt to predict the muon’s scattering points (or full trajectory) using a neural network using simulated data and use these as an input to the model.

8 Conclusion

In conclusion, we have constructed a state-of-the-art model for muon scattering tomography which outperforms traditional methods such as PoCA and MLEM. Furthermore, we find that our model is robust to various corruptions, with its performance barely changing when they are applied. We hope that our research will spark further investigation into the usage of deep learning in this field. Improvements in imaging techniques for muon scattering tomography will have wide-ranging applications from ensuring nuclear non-proliferation to the discovery of secret chambers in ancient structures.

Declaration of Past Submission

This project was previously submitted for the Beamline for Schools (BL4S) 2022.

References

- [1] Valeri Tioukov, Andrey Alexandrov, Cristiano Bozza, Lucia Consiglio, Nicola D'Ambrosio, Giovanni De Lellis, Chiara De Sio, Flora Giudicepietro, Giovanni Macedonio, Seigo Miyamoto, Ryuichi Nishiyama, Massimo Orazi, Rosario Peluso, Andrey Sheshukov, Chiara Sirignano, Simona Maria Stellacci, Paolo Strolin, and Hiroyuki K. M. Tanaka. First muography of stromboli volcano. *Scientific Reports*, 9(1), April 2019.
- [2] G. Saracino, F. Ambrosino, L. Bonechi, L. Cimmino, R. D'Alessandro, M. D'Errico, P. Noli, L. Scognamiglio, and P. Strolin. Applications of muon absorption radiography to the fields of archaeology and civil engineering. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 377(2137):20180057, December 2018.
- [3] Diletta Borselli, Tommaso Beni, Lorenzo Bonechi, Massimo Bongi, Debora Brocchini, Nicola Casagli, Roberto Ciaranfi, Luigi Cimmino, Vitaliano Ciulli, Raffaello D'Alessandro, Andrea Dini, Catalin Frosin, Giovanni Gigli, Sandro Gonzi, Silvia Guideri, Luca Lombardi, Massimiliano Nocentini, and Giulio Saracino. Three-dimensional muon imaging of cavities inside the temperino mine (italy). *Scientific Reports*, 12(1), December 2022.
- [4] Sébastien Procureur, Kunihiro Morishima, Mitsuaki Kuno, Yuta Manabe, Nobuko Kitagawa, Akira Nishio, Hector Gomez, David Attié, Ami Sakakibara, Kotaro Hikata, Masaki Moto, Irakli Mandjavidze, Patrick Magnier, Marion Lehuraux, Théophile Benoit, Denis Calvet, Xavier Coppolani, Mariam Kebbiri, Philippe Mas, Hany Helal, Mehdi Tayoubi, Benoit Marini, Nicolas Serikoff, Hamada Anwar, Vincent Steiger, Fumihiko Takasaki, Hirofumi Fujii, Kotaro Satoh, Hideyo Kodama, Kohei Hayashi, Pierre Gable, Emmanuel Guerriero, Jean-Baptiste Mouret, Tamer Elnady, Yasser Elshayeb, and Mohamed Elkarmoty. Precise characterization of a corridor-shaped structure in khufu's pyramid by observation of cosmic-ray muons. *Nature Communications*, 14(1), March 2023.
- [5] J. Beringer, J. F. Arguin, R. M. Barnett, K. Copic, O. Dahl, D. E. Groom, C. J. Lin, J. Lys, H. Murayama, C. G. Wohl, W. M. Yao, P. A. Zyla, C. Amsler, M. Antonelli, D. M. Asner, H. Baer, H. R. Band, T. Basaglia, C. W. Bauer, J. J. Beatty, V. I. Belousov, E. Bergren, G. Bernardi, W. Bertl, S. Bethke, H. Bichsel, O. Biebel, E. Blucher, S. Blusk, G. Brooijmans, O. Buchmueller, R. N. Cahn, M. Carena, A. Ceccucci, D. Chakraborty, M. C. Chen, R. S. Chivukula, G. Cowan, G. D'Ambrosio, T. Damour, D. de Florian, A. de Gouvêa, T. DeGrand, P. de Jong, G. Dissertori, B. Dobrescu, M. Doser, M. Drees, D. A. Edwards, S. Eidelman, J. Erler, V. V. Ezhela, W. Fetscher, B. D. Fields, B. Foster, T. K. Gaisser, L. Garren, H. J. Gerber, G. Gerbier, T. Gherghetta, S. Golwala, M. Goodman, C. Grab, A. V. Gritsan,

J. F. Grivaz, M. Grünewald, A. Gurtu, T. Gutsche, H. E. Haber, K. Hagiwara, C. Hagmann, C. Hanhart, S. Hashimoto, K. G. Hayes, M. Heffner, B. Heltsley, J. J. Hernández-Rey, K. Hikasa, A. Höcker, J. Holder, A. Holtkamp, J. Huston, J. D. Jackson, K. F. Johnson, T. Junk, D. Karlen, D. Kirkby, S. R. Klein, E. Klempt, R. V. Kowalewski, F. Krauss, M. Kreps, B. Krusche, Yu. V. Kuyanov, Y. Kwon, O. Lahav, J. Laiho, P. Langacker, A. Liddle, Z. Ligeti, T. M. Liss, L. Littenberg, K. S. Lugovsky, S. B. Lugovsky, T. Mannel, A. V. Manohar, W. J. Marciano, A. D. Martin, A. Masoni, J. Matthews, D. Milstead, R. Miquel, K. Mönig, F. Moortgat, K. Nakamura, M. Narain, P. Nason, S. Navas, M. Neubert, P. Nevski, Y. Nir, K. A. Olive, L. Pape, J. Parsons, C. Patrignani, J. A. Peacock, S. T. Petcov, A. Piepke, A. Pomarol, G. Punzi, A. Quadt, S. Raby, G. Raffelt, B. N. Ratcliff, P. Richardson, S. Roesler, S. Rolli, A. Romaniouk, L. J. Rosenberg, J. L. Rosner, C. T. Sachrajda, Y. Sakai, G. P. Salam, S. Sarkar, F. Sauli, O. Schneider, K. Scholberg, D. Scott, W. G. Seligman, M. H. Shaevitz, S. R. Sharpe, M. Silari, T. Sjöstrand, P. Skands, J. G. Smith, G. F. Smoot, S. Spanier, H. Spieler, A. Stahl, T. Stanev, S. L. Stone, T. Sumiyoshi, M. J. Syphers, F. Takahashi, M. Tanabashi, J. Terning, M. Titov, N. P. Tkachenko, N. A. Törnqvist, D. Tovey, G. Valencia, K. van Bibber, G. Venanzoni, M. G. Vinciter, P. Vogel, A. Vogt, W. Walkowiak, C. W. Walter, D. R. Ward, T. Watari, G. Weiglein, E. J. Weinberg, L. R. Wiencke, L. Wolfenstein, J. Womersley, C. L. Woody, R. L. Workman, A. Yamamoto, G. P. Zeller, O. V. Zenin, J. Zhang, R. Y. Zhu, G. Harper, V. S. Lugovsky, and P. Schaffner. Review of particle physics. *Physical Review D*, 86(1), July 2012.

- [6] Larry Joe Schultz. *Cosmic Ray Muon Radiography*. PhD thesis, Portland State University, 2003.
- [7] Larry J Schultz, Gary S Blanpied, Konstantin N Borozdin, Andrew M Fraser, Nicolas W Hengartner, Alexei V Klimenko, Christopher L Morris, Chris Orum, and Michael J Sossong. Statistical reconstruction for cosmic ray muon tomography. *IEEE transactions on Image Processing*, 16(8):1985–1993, 2007.
- [8] Christopher J. Benton, Nathan D. Smith, Stephen J. Quillin, and Christopher A. Steer. Most probable trajectory of a muon in a scattering medium, when input and output trajectories are known. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 693:154–159, November 2012.
- [9] Guobao Wang, L.J. Schultz, and Jinyi Qi. Bayesian image reconstruction for improving detection performance of muon tomography. *IEEE Transactions on Image Processing*, 18(5):1080–1089, May 2009.
- [10] R. W. Schulte, S. N. Penfold, J. T. Tafas, and K. E. Schubert. A maximum likelihood proton path formalism for application in proton computed tomography. *Medical Physics*, 35(11):4849–4856, oct 2008.
- [11] Hengguan Yi, Zhi Zeng, Baihui Yu, Jianping Cheng, Ziran Zhao, Xuewu Wang, Ming Zeng, and Yi Wang. Bayesian-theory-based most probable trajectory reconstruction algorithm in cosmic ray muon tomography. In *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–4, 2014.
- [12] Stylianos Chatzidakis, Zhengzhi Liu, Jason P. Hayward, and John M. Scaglione. A generalized muon trajectory estimation algorithm with energy loss for application to muon tomography. *Journal of Applied Physics*, 123(12), mar 2018.
- [13] G. Jonkmans, V.N.P. Anghel, C. Jewett, and M. Thompson. Nuclear waste imaging and spent fuel verification by muon tomography. *Annals of Nuclear Energy*, 53:267–273, March 2013.

- [14] M. Stapleton, J. Burns, S. Quillin, and C. Steer. Angle statistics reconstruction: a robust reconstruction algorithm for muon scattering tomography. *Journal of Instrumentation*, 9(11):P11019–P11019, November 2014.
- [15] C. Thomay, P. Baesso, D. Cussans, J. Davies, P. Glaysher, S. Quillin, S. Robertson, C. Steer, C. Vassallo, and J. Velthuis. A novel technique to detect special nuclear material using cosmic rays. *Geoscientific Instrumentation, Methods and Data Systems*, 1(2):235–238, December 2012.
- [16] Linjun Hou, Quanhu Zhang, Jianqing Yang, Xingfu Cai, Qingxu Yao, Yonggang Huo, and Qifan Chen. A novel reconstruction algorithm based on density clustering for cosmic-ray muon scattering inspection. *Nuclear Engineering and Technology*, 53(7):2348–2356, July 2021.
- [17] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytrcek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J.J. Gómez Cadenas, I. González, G. Gracia Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F.W. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampén, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. Mora de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M.G. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. Di Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. Safai Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander, H. Weber, J.P. Wellisch, T. Wenaus, D.C. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschesche. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, July 2003.
- [18] Timothy P. Szczykutowicz, Giuseppe V. Toia, Amar Dhanantwari, and Brian Nett. A review of deep learning CT reconstruction: Concepts, limitations, and promise in clinical practice. *Current Radiology Reports*, 10(9):101–115, July 2022.
- [19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells III, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer, 2015.
- [20] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 11966–11976. IEEE, 2022.
- [21] Bruno Rossi and Kenneth Greisen. Cosmic-ray theory. *Reviews of Modern Physics*, 13(4):240, 1941.
- [22] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances*

- in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008, 2017.
- [23] Saifullahi Aminu Bello, Shangshu Yu, Cheng Wang, Jibril Muhmmad Adam, and Jonathan Li. Deep learning on 3d point clouds. *Remote Sensing*, 12(11):1729, 2020.
- [24] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.
- [25] Ngoc-Vuong Ho, Tan Nguyen, Gia-Han Diep, Ngan Le, and Binh-Son Hua. Point-unet: A context-aware point-based neural network for volumetric segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 644–655. Springer, 2021.
- [26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016.
- [27] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.
- [28] Weihe Zeng, M Zeng, X Pan, Z Zeng, H Ma, and Jianping Cheng. Principle study of image reconstruction algorithms in muon tomography. *Journal of Instrumentation*, 15(02):T02005, 2020.
- [29] Prashant Shukla and Sundaresh Sankrith. Energy and angular distributions of atmospheric muons at the earth. *International Journal of Modern Physics A*, 33(30):1850175, October 2018.
- [30] Sébastien Procureur, Kunihiko Morishima, Mitsuaki Kuno, Yuta Manabe, Nobuko Kitagawa, Akira Nishio, Hector Gomez, David Attié, Ami Sakakibara, Kotaro Hikata, Masaki Moto, Irakli Mandjavidze, Patrick Magnier, Marion Lehuraux, Théophile Benoit, Denis Calvet, Xavier Coppolani, Mariam Kebbiri, Philippe Mas, Hany Helal, Mehdi Tayoubi, Benoit Marini, Nicolas Serikoff, Hamada Anwar, Vincent Steiger, Fumihiko Takasaki, Hirofumi Fujii, Kotaro Satoh, Hideyo Kodama, Kohei Hayashi, Pierre Gable, Emmanuel Guerriero, Jean-Baptiste Mouret, Tamer Elnady, Yasser Elshayeb, and Mohamed Elkarmoty. Precise characterization of a corridor-shaped structure in khufu’s pyramid by observation of cosmic-ray muons. *Nature Communications*, 14(1), March 2023.
- [31] Luigi Cimmino, Guglielmo Baccani, Pasquale Noli, Lucio Amato, Fabio Ambrosino, Lorenzo Bonechi, Massimo Bongi, Vitaliano Ciulli, Raffaello D’Alessandro, Mariaelena D’Errico, Sandro Gonzi, Barbara Melon, Gianluca Minin, Giulio Saracino, Luca Scognamiglio, Paolo Strolin, and Lorenzo Viliiani. 3d muography for the search of hidden cavities. *Scientific Reports*, 9(1), February 2019.
- [32] G. Saracino, L. Amato, F. Ambrosino, G. Antonucci, L. Bonechi, L. Cimmino, L. Consiglio, R. D’Alessandro, E. De Luzio, G. Minin, P. Noli, L. Scognamiglio, P. Strolin, and A. Varriale. Imaging of underground cavities with cosmic-ray muons from observations at mt. echia (naples). *Scientific Reports*, 7(1), April 2017.
- [33] Valeri Tioukov, Kunihiko Morishima, Carlo Leggieri, Federico Caprioli, Nobuko Kitagawa, Mitsuaki Kuno, Yuta Manabe, Akira Nishio, Andrey Alexandrov, Valerio Gentile, Antonio Iuliano, and Giovanni De Lellis. Hidden chamber discovery in the underground hellenistic necropolis of neapolis by muography. *Scientific Reports*, 13(1), April 2023.

- [34] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*, volume 9906 of *Lecture Notes in Computer Science*, pages 694–711. Springer, 2016.
- [35] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6):1348–1357, 2018.
- [36] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Appendices

Appendix A Proofs

To prove [Theorem 1](#), we shall split the theorem into the arbitrarily large resolution case and the arbitrarily large point size case.

A.1 Arbitrarily Large Resolution

Theorem 2. Suppose $f : \chi \rightarrow \mathbb{R}^p$ is a continuous set function w.r.t $d_H(\cdot, \cdot)$, such that for all $\epsilon > 0$, there exists some configuration of the model parameters θ for sufficiently large p , such for that any $S \in \chi$ ²

$$\left| f(S) - \gamma_\theta \left(\frac{\sum_{x_i \in S} \{\phi(\eta(x_i)) \cdot h_\theta(x_i)\}}{\sum_{x_i \in S} \{\phi(\eta(x_i)) \cdot J_{d \times c}\}} \right) \right| < \epsilon$$

where $\gamma_\theta : \mathbb{R}^{p \times c} \rightarrow \mathbb{R}^p$ is any continuous function, $h_\theta : \mathbb{R}^m \rightarrow \mathbb{R}^{d \times c}$ is any continuous function, $\eta : \mathbb{R}^m \rightarrow \mathbb{R}$, $\phi : \mathbb{R}^m \rightarrow \mathbb{R}^{p \times d}$ and $J_{d \times c}$ is the ones matrix of shape (d, c) . η represents the PoCA function that generates a scattering point from the muon detection. ϕ is an indicator function for a set of intervals of length d derived from its input. The indicator function for each of these intervals is placed along one row in the last dimension. γ_θ and h_θ can be taken to be any continuous function due to the universal approximation theorem for CNNs and MLPs.

Proof. The idea is that for a sufficiently large p , the indicator functions of ϕ will not overlap, allowing S to be recovered exactly using an inverse function T .

Let h_θ simply be the identity function.³ Now, consider a function $\mathcal{T} : \mathbb{R}^{p \times m} \rightarrow \chi$, $\mathcal{T}(X) = \{x : |x| > 0, x \in \mathbb{R}^m, x \text{ is an entry in the last dimension of } X\}$. Now, we define γ_θ as $f \circ \mathcal{T}$. Clearly,

$$\left| f(S) - f \left(\mathcal{T} \left(\frac{\sum_{x_i \in S} \{\phi(\eta(x_i)) \cdot h_\theta(x_i)\}}{\sum_{x_i \in S} \{\phi(\eta(x_i)) \cdot J_{d \times c}\}} \right) \right) \right| = |f(S) - f(S)| = 0 < \epsilon$$

■

A.2 Arbitrarily Large Point Size

Theorem 3. Suppose $f : \chi \rightarrow \mathbb{R}^p$ is a continuous set function w.r.t $d_H(\cdot, \cdot)$, such that for all $\epsilon > 0$, there exists some configuration of the model parameters θ for $\phi(\eta(x_i)) = J_{p \times d}$, such that for any $S \in \chi$,

$$\left| f(S) - \gamma_\theta \left(\frac{\sum_{x_i \in S} \{\phi(\eta(x_i)) \cdot h_\theta(x_i)\}}{\sum_{x_i \in S} \{\phi(\eta(x_i)) \cdot J_{d \times c}\}} \right) \right| < \epsilon$$

where $\gamma_\theta : \mathbb{R}^{p \times c} \rightarrow \mathbb{R}^p$ is any continuous function, $h_\theta : \mathbb{R}^m \rightarrow \mathbb{R}^{d \times c}$ is any continuous function, $\eta : \mathbb{R}^m \rightarrow \mathbb{R}$, $\phi : \mathbb{R}^m \rightarrow \mathbb{R}^{p \times d}$ and $J_{d \times c}$ is the ones matrix of shape (d, c) . η represents the PoCA function that generates a scattering point from the muon detection. ϕ is an indicator function for a set of intervals of length d derived from its input. The indicator function for each of these intervals is placed along one row in

²For this division, we assume that the division is element-wise and if division by 0 occurs the value will just be 0.

³Since h_θ is the identity function, $c = m$

the last dimension. γ_θ and h_θ can be taken to be any continuous function due to the universal approximation theorem for CNNs and MLPs.

Proof. In the case of $\phi(\eta(x_i)) = J_{p \times d}$, the theorem reduces to

$$\left| f(S) - \gamma_\theta \left(\frac{\sum_{x_i \in S} \{h_\theta(x_i)\}}{N} \right) \right| < \epsilon$$

where N is the number of muons. By invoking Theorem 7 in [36], any continuous permutation invariant set function can be decomposed into the form

$$f(S) = \rho \left(\sum_{x_i \in S} \{\psi(x_i)\} \right)$$

where ρ and ψ are continuous functions. By the universal approximation theorem for CNNs and MLPs, there exists some θ such that

$$\rho \left(\sum_{x_i \in S} \psi(x_i) \right) = \gamma_\theta \left(\frac{\sum_{x_i \in S} \{h_\theta(x_i)\}}{N} \right)$$

Therefore

$$\left| f(S) - \gamma_\theta \left(\frac{\sum_{x_i \in S} \{h_\theta(x_i)\}}{N} \right) \right| = |f(S) - f(S)| < \epsilon$$

■

Appendix B Raw Results

Table 3: Results of various models for different amounts of dosage

Model	Dosage	MSE↓	MAE↓	PSNR↑
μ -Net	1024	0.2674	0.2245	16.4496
PoCA	1024	0.4191	-	14.4066
μ -Net	2048	0.2009	0.1883	17.7133
PoCA	2048	0.3723	-	14.9207
μ -Net	4096	0.1701	0.1698	18.4348
PoCA	4096	0.3187	-	15.5964
μ -Net	8192	0.1939	0.1486	19.1594
PoCA	8192	0.2861	-	16.0640
μ -Net	16384	0.1321	0.1269	19.6145
PoCA	16384	0.2701	-	16.3143

Table 4: Results of various methods for different levels of error in the momentum estimate. μ -Net* indicates that the model was finetuned on the new data for 10 epochs. The best results are bolded.

Model	Detector Resolution	MSE↓	MAE↓	PSNR↑
μ -Net	64×64	0.2355	0.2197	16.9981
μ -Net*	64×64	0.2233	0.2090	17.2569
PoCA	64×64	0.5210	0.2690	13.1784
μ -Net	128×128	0.2114	0.1978	17.4785
μ -Net*	128×128	0.2077	0.2015	17.5769
PoCA	128×128	0.5226	0.2696	13.1625
μ -Net	256×256	0.2043	0.1914	17.6305
μ -Net*	256×256	0.2039	0.1988	17.6426
PoCA	256×256	0.5210	0.2690	13.1784
μ -Net	512×512	0.2012	0.1885	17.7043
μ -Net*	512×512	0.2014	0.2011	17.7025
PoCA	512×512	0.5210	0.2690	13.1784
μ -Net	1024×1024	0.2005	0.1883	17.7176
μ -Net*	1024×1024	0.2023	0.1914	17.7003
PoCA	1024×1024	0.5210	0.2690	13.1784
μ -Net	2048×2048	0.2005	0.1881	17.7120
μ -Net*	2048×2048	0.1993	0.1996	17.7592
PoCA	2048×2048	0.5210	0.2690	13.1784
μ -Net	∞	0.2009	0.1883	17.7133
PoCA	∞	0.3726	0.2443	14.6382

Table 5: Results of various models for different levels of error in the momentum estimate. μ -Net* indicates that the model was finetuned on the new data for 10 epochs. The best results are bolded.

Model	Δp	MSE↓	MAE↓	PSNR↑
μ -Net	0%	0.1997	0.1871	17.7392
μ -Net*	0%	0.1941	0.1899	17.8689
PoCA	0%	0.3759	-	14.8927
μ -Net	10%	0.1995	0.1876	17.7349
μ -Net*	10%	0.1966	0.1861	17.8051
PoCA	10%	0.3843	-	14.7828
μ -Net	20%	0.2009	0.1883	17.7133
PoCA	20%	0.3726	0.2443	14.6382
μ -Net	40%	0.2034	0.1900	17.6543
μ -Net*	40%	0.2018	0.1908	17.6872
PoCA	40%	0.5709	-	14.9207
μ -Net	80%	0.2214	0.2021	17.2452
μ -Net*	80%	0.2033	0.2009	17.6690
PoCA	80%	1.2158	-	13.0640
μ -Net	100%	0.2308	0.2080	17.0444
μ -Net*	100%	0.2050	0.2073	17.6139
PoCA	100%	1.9289	0.3697	9.028
μ -Net	∞	0.5131	0.3388	13.0599
μ -Net*	∞	0.2048	0.1942	17.6251
PoCA	∞	1.8819	0.3683	9.3671